Yangon University of Economics
Post Graduate Diploma in Research Studies (PGDRS) Progamme
Third Quarter Examination, March 2024

PGDRS-313
PGDRS (9th Batch) **Computer Aided Statistical Analysis**
Answer **All** Questions. **Time Allowed: 3 hours**

1. (a) A company manager wants to know the performance of employees in a company. The manager collected the performance of 42 sample employees. The data is as follows.

32	33	31	34	29	36	27	31	35	36	30	29	33	32
35	31	25	22	31	32	25	36	31	30	30	29	20	18
38	30	35	27	31	27	32	32	34	34	37	31	30	20

(i) Calculate the measure of central tendency and interpret the results.
(ii) Calculate the measure of variation and interpret the results.
(iii) Calculate the measure of location in terms of 40th, 65th, 90th percentiles, quartiles, and interquartile range.
(iv) Find skewness and kurtosis and comment on your results.
(v) Analyzing histogram, normal Q-Q plot, and tests of normality, what is your conclusion regarding normality? If a transformation is necessary, which one would you use?

(b) The ages at the time of death of those Vice Presidents of the United States who have passed away are listed below.

90	83	80	73	70	51	68	79	70	71
72	74	67	54	81	66	62	63	68	57
66	96	78	55	60	66	57	71	60	85

(i) Check whether the given data is normal or not.
(ii) Check the data set for outliers.
(iii) Test whether the mean age at the time of death of those Vice Presidents of the United States who have passed away is 63 years old.

2. (a) State the assumptions of the independent sample t test.

(b) The number of stories in two randomly selected samples of tall buildings in Miami and Houston are shown.

Miami	Houston
64,57,52,45,60,55,47,51,56,52,60,50,49, 48,56,63,43,41,53,49, 58,34,45,44, 46	71,53,55,47,46,64,55,40,36,36,54,34,52,45, 35,49,41,40,37,33,50,42,40,37,33

(i) Construct a 95% confidence interval for the difference in average stories of buildings in two cities.
(ii) Is there difference in variance between stories of buildings in two cities?
(iii) Test the hypothesis whether the average stories of buildings in Miami is more than average stories of buildings in Houston at 5% level of significance.

P. T. O/ 2

3. (a) Discuss the assumptions of Chi-square test of independence.

(b) A survey conducted 194 women based on educational level, marital status and having knowledge about gender violence. By using the collected dataset ***(demographic data.xlsx):***

(i) Test whether having knowledge about gender violence and marital status are independent or not. ($\alpha=0.05$)

(ii) Test whether having knowledge about gender violence and education level are independent or not. ($\alpha=0.05$)

4. (a) State what values you would assign to dummy variables to measure a person's race if the categories included (1) white, (2) black, (3) oriental, and (4) other.

(b) Owners of Mightly Muscles are trying to increase membership at the club. After reviewing operations at other health clubs in the area, Harry Hunk, the manager at Mighty Muscles, feels that there might be a relationship between the number of members in a health club and several explanatory variables, including (1) monthly dues, (2) local population, and (3) type of weight-training equipment in the club. Three kinds of weight machines are commonly found in health clubs: MuscleUp (MU), Iron Man (IM), and Studly Dude (SD). Mr. Hunk collects data for these variables for 20 health clubs similar to Mighty Muscles. Analyze the data and determine how well the model fits the data. Interpret your findings and recommend to Harry the policy decisions he should make for the club.

Membership	Dues	Population (in 1,000s)	Equipment Type
345	$22.00	39.7	MU
387	21	43.3	MU
412	20	47.7	MU
434	20	48.9	SD
467	18.5	54.5	SD
490	18	55.7	SD
612	15	62.1	SD
523	17.5	56.7	IM
545	17	58.9	IM
567	16.5	60	IM
578	16	61.3	SD
599	15	63.6	SD
634	14.5	60	IM
656	13	61	SD
669	13.1	61.4	MU
134	30	23.5	MU
243	27	27.8	IM
259	25.5	32.1	IM
278	23	37.3	SD
321	24	35.8	SD

P. T. O/ 2

(i) Determine the multiple regression equation and interpret the regression coefficients.
(ii) What is the value of adjusted R^2? Comment on this value.
(iii) Conduct a test of hypothesis to determine whether the relationship between the number of members in a health club and all independent variables.
(iv) Conduct a test of hypothesis to determine whether the relationship between the number of members in a health club and monthly dues.
(v) Is there a multicollinearity problem between independent variables?

5. (a) Describe the assumptions of one-way ANOVA.

(b) A city council wants to assess income disparities across three different neighborhoods (Downtown, Suburbia, and Industrial). They collected annual income data (in thousands of dollars) from a random sample of residents in each neighborhood. The data is as follows.

Downtown: 60, 65, 70, 55, 62, 58, 63, 59, 64, 68
Suburbia: 75, 80, 78, 82, 85, 79, 81, 77, 76, 83
Industrial: 45, 48, 50, 42, 47, 43, 49, 46, 52, 44

(i) Assess whether the assumptions of normality and homogeneity of variance are met on this data.
(ii) Test whether there are significant differences in mean income among three neighborhoods. ($\alpha = 0.05$)
(iii) Conduct post hoc tests to identify which neighborhoods have significantly different income levels.
(iv) Calculate a 95% confidence interval for difference mean for each pair.